Competing Against Nash Equilibria in Adversarially Changing Zero-Sum Games

Adrian Rivera Cardoso

Joint work with: Jacob Abernethy, He Wang, Huan Xu

Georgia Institute of Technology

June 7, 2019

Overview

1. Matrix Games
2. Online Matrix Games
3. An Impossibility Result
4. Good News

Matrix Games

One of the canonical problems in game theory are **zero-sum matrix games**. Finding a Nash Equilibrium is core to many problems in statistics, optimization and economics.

Setup:

- Player 1 chooses a probability distribution x over d_1 actions
- Player 2 chooses a probability distribution y over d_2 actions
- The payoffs are specified by matrix $A \in \mathbb{R}^{d_1 \times d_2}$.
- A_{ij} encodes the loss of Player 1 $=$ reward of Player 2 when they play actions i, j respectively

Matrix Games

The goal is to find a Nash Equilibrium (NE) of the game. A NE is a pair (x^*, y^*) such that for all $x \in \Delta^{d_1}$, $y \in \Delta^{d_2}$ it holds that

$$x^{*\top} Ay \leq x^{*\top} Ay^* \leq x^\top Ay^*$$

$(x^*)^\top Ay^*$ is called the value of the game an it holds that

$$x^{*\top} Ay^* = \min_{x \in \Delta^{d_1}} \max_{y \in \Delta^{d_2}} x^\top Ay = \max_{y \in \Delta^{d_2}} \min_{x \in \Delta^{d_1}} x^\top Ay.$$

How to find a NE? Run two OCO algorithms in parallel and then average the history of iterates.

But what if the payoff matrix changes with time?!?

Online Matrix Games: Problem Setup

- Two players play a sequence of Matrix Games for T time steps
- In step t they must each choose a distribution over actions $x_t \in \Delta^{d_1}$, $y_t \in \Delta^{d_2}$
- An adversary chooses payoff matrix A_t
- They receive loss/reward equal to $x_t^\top A_t y_t$ and observe A_t
- Using this new information they choose x_{t+1}, y_{t+1}

Their goal is to achieve sublinear Nash Equilibrium Regret

$$NE.Regret \triangleq \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta^{d_1}} \max_{y \in \Delta^{d_2}} \sum_{t=1}^{T} x^\top A_t y \right|$$

An Impossibility Result

We know that when $A_t = A$ for all $t = 1, ..., T$, if each player minimizes its own Individual Regret,

$$\sum_{t=1}^{T} f_t(x_t) - \min_{x \in X} \sum_{t=1}^{T} f_t(x),$$

and we average their iterates we find a NE equilibrium.

Is this still a good strategy to minimize Individual Regret when $A_t \neq A$ for all $t = 1, ..., T$?

An Impossibility Result

Theorem

Consider any algorithm that selects a sequence of x_t, y_t pairs given the past payoff matrices $A_1, \ldots, A_{t-1}$. Consider the following three objectives:

$$
\begin{aligned}
\left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta^{d_1}} \max_{y_t \in \Delta^{d_2}} \sum_{t=1}^{T} x^\top A_t y \right| &= o(T), && (1)\\
\sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_X} \sum_{t=1}^{T} x^\top A_t y_t &= o(T), && (2)\\
\max_{y \in \Delta_Y} \sum_{t=1}^{T} x_t^\top A_t y - \sum_{t=1}^{T} x_t^\top A_t y_t &= o(T). && (3)
\end{aligned}
$$

Then there exists an (adversarially-chosen) sequence $A_1, A_2, \ldots$ such that not all of (1), (2), *and* (3), *are true.*

Good News

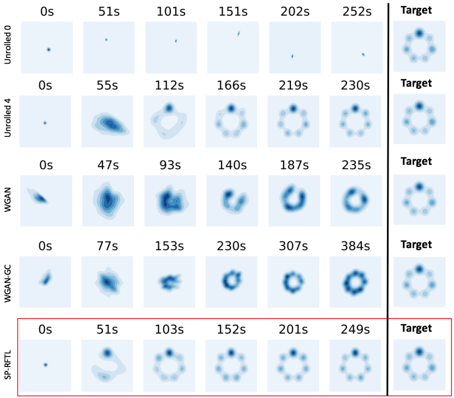
Theorem

There exists an algorithm (see paper or poster) that guarantees:

$$NE.Regret \leq O(\sqrt{T}\ln(T) + \max\{\ln(d_1), \ln(d_2)\}\sqrt{T})$$

Some Preliminary Results

Our algorithm seems to be useful for training GANs.



Different algorithms used for training GANs on the mixture of Gaussians data set.

Thank you!

See you at Pacific Ballroom 151 from 6:30-9:00 pm